Genomics 99 (2012) 347–354

Contents lists available at SciVerse ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Comparison of pathways associated with hepatitis B- and C-infected hepatocellular carcinoma using pathway-based class discrimination method

Sun Young Lee [a,b], Kwang Hoon Song [a,c], Imhoi Koo [d], Kee-Ho Lee [e], Kyung-Suk Suh [f], Bu-Yeo Kim [a,c,*]

[a] Division of Constitutional Medicine Research, Korea Institute of Oriental Medicine, Daejeon, Republic of Korea
[b] School of Computational Sciences and Center for In Silico Protein Science, Korea Institute for Advanced Study, Seoul, Republic of Korea
[c] Herbal Medicine Research Division, Korea Institute of Oriental Medicine, Daejeon, Republic of Korea
[d] Department of Bioinformatics and Biostatistics, University of Louisville, Louisville, USA
[e] Laboratory of Radiation Molecular Oncology, Korea Institute of Radiological and Medical Sciences, Seoul, Republic of Korea
[f] Department of Surgery, Seoul National University School of Medicine, Seoul, Republic of Korea

ARTICLE INFO

Article history:
Received 17 November 2011
Accepted 20 April 2012
Available online 29 April 2012

Keywords:
Hepatocellular carcinoma
Pathway
Microarray
HBV
HCV
Prediction

ABSTRACT

Molecular signatures causing hepatocellular carcinoma (HCC) from chronic infection of hepatitis B virus (HBV) or hepatitis C virus (HCV) are not clearly known. Using microarray datasets composed of HCV-positive HCC or HBV-positive HCC, pathways that could discriminate tumor tissue from adjacent non-tumor liver tissue were selected by implementing nearest shrunken centroid algorithm. Cancer-related signaling pathways and lipid metabolism-related pathways were predominantly enriched in HCV-positive HCC, whereas functionally diverse pathways including immune-related pathways, cell cycle pathways, and RNA metabolism pathways were mainly enriched in HBV-positive HCC. In addition to differentially involved pathways, signaling pathways such as TGF-β, MAPK, and p53 pathways were commonly significant in both HCCs, suggesting the presence of common hepatocarcinogenesis process. The pathway clustering also verified segregation of pathways into the functional subgroups in both HCCs. This study indicates the functional distinction and similarity on the pathways implicated in the development of HCV- and/or HBV-positive HCC.

1. Introduction

Hepatocellular carcinoma (HCC) is one of the most fatal cancers worldwide, and about half million patients die from this disease each year [1]. Chronic infection of hepatitis B virus (HBV) or hepatitis C virus (HCV) is the major risk factors for HCC. Although histological evaluation and clinical manifestation are indistinguishable between chronic hepatitis patients with HBV or HCV infection [2], gene expression pattern in HBV- or HCV-infected livers has been reported to be different [3,4]. Moreover, microarray studies have demonstrated that HBV and HCV cause hepatocarcinogenesis by different molecular mechanisms [5–8]. For example, based on the differentially expressed genes between HBV- and HCV-positive HCC, biological functions such as apoptosis, DNA repair responses, and inflammatory pathways have been differently implicated in two types of HCC [4,5]. Although the microarray revealed the different molecular signatures responsible for development of HBV-positive HCC or HCV-positive HCC, the different microarray platforms and heterogeneous clinicopathological nature of samples used in the experiments makes it hard to obtain common genes or common signaling pathways that can be applied to all datasets. For example, we previously reported that only a small part of the principle genes in HBV-positive HCC were in common with those selected from other microarray datasets [9,10]. Moreover, since most research involving high-throughput data have eventually focused on only a handful of significant genes, it is difficult to obtain biological information that can be extracted from a network relationship of whole genes. In addition, individual genes are more susceptible to noise inherent to the microarray technology. One of the approaches to overcome these limitations is using functionally- or structurally-related pre-defined gene sets, such as pathway, rather than using individual genes [11,12]. Recent reports also demonstrated that the pathway-based approach yields more interpretable results on particular cellular or physiological functions by simplifying the complex structure of a genetic network [13–15]. A variety of methods have been proposed to identify pathways associated with the phenotype of subjects, including discriminant analysis [16,17] and enrichment methods [11,18]. Recently, Pang et al. used random forest algorithm to identify

Abbreviations: HCC, Hepatocellular carcinoma; HBV, Hepatitis B virus; HCV, Hepatitis C virus; PAM, Prediction Analysis of Microarrays; RF, Random forest; SVM, Support vector machine; LDA, Linear discriminant analysis; KNN, k-nearest neighbor classifier; GSA, Gene set analysis; CV, Cross-validation; FDR, False discovery rate; MAPK, Mitogen-activated protein kinase; TGF-β, Transforming growth factor-beta; JAK/STAT, Janus kinase/signal transducer and activator of transcription; SKP2, S-phase kinase-associated protein 2.

* Corresponding author at: Herbal Medicine Research Division, Korea Institute of Oriental Medicine, 1672 Yuseongdae-ro, Yuseong-gu, Daejeon, 305-811, Republic of Korea. Fax: +82 42 868 9480.
E-mail address: buykim@kiom.re.kr (B.-Y. Kim).

0888-7543/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygeno.2012.04.004

pathways discriminating sample classes and further built pathway clusters to understand possible crosstalk between pathways [19,20].

Because biological processes associated with development of HCC have not been clearly elucidated in pathway level, it is important to identify the functional changes involved in the development of HBV-positive HCC or HCV-positive HCC. In this study, we identified that cancer-related signaling pathways and lipid metabolism-related pathways were predominantly enriched in HCV-positive HCC, whereas diverse functions including immune-related pathways, cell cycle pathways, and RNA metabolism pathways were mainly enriched in HBV-positive HCC. In addition, many cancer-related signaling pathways were also commonly significant in both HBV-positive HCC and HCV-positive HCC indicating the presence of common hepatocarcinogenesis mechanism.

2. Results

2.1. Assessment of the discrimination methods

The overall analysis procedure is depicted in Fig. 1, in which diverse classification algorithms were implemented to identify pathways using the random training/test method. We measured the efficiency of five different models (PAM, RF, SVM, KNN and LDA) on class discrimination, namely, tumor versus adjacent non-tumor liver tissue, using three different HCC microarray datasets: HCV-positive dataset, HBV-positive dataset and HBV- or HCV-positive dataset. Clinicopathological information about the three datasets is shown in Supplementary Table 1. Fig. 2 shows the correlation plot of averaged test error rate for each pathway according to classification methods on the HCV-positive HCC dataset. PAM, RF, and SVM displayed similar distribution of averaged test error rate with correlation coefficient exceeding 0.6. The scatter plot matrixes for the HBV-positive HCC dataset and HBV- or HCV-positive HCC dataset also showed similar results (Supplementary Figs. 1 and 2). We selected PAM as the main discrimination method for the following analysis since this algorithm was reported to be excellent in microarray analysis [9,21].

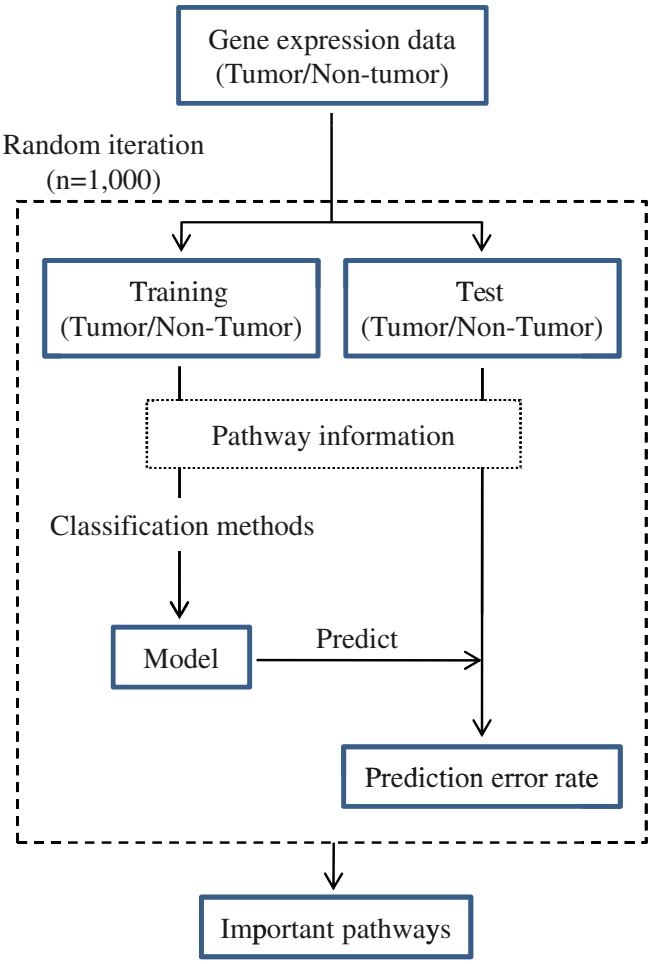

Fig. 1. A schematic illustration of the analysis strategy. Initial dataset consisting of HCC and adjacent non-tumor liver tissue was randomly divided into training and test sets. Using genes included in each pathway, a training model discriminating tumor tissue from adjacent non-tumor liver tissue was built by implementing various classification algorithms: nearest shrunken centroid (or prediction analysis of microarrays; PAM), random forest (RF), support vector machine (SVM), linear discriminant analysis (LDA) and k-nearest neighbor (KNN) classifier. Then, the model of each pathway was evaluated in the test set. This procedure was repeated 1000 times with random training and test sets. Finally, the averaged test error rate was measured to select significant pathways.

2.2. Identification of pathways

By implementing PAM, we identified pathways implicated in liver tumorigenesis and compared these pathways with those from the GSA algorithm, a recently-reported pathway-based method [18]. Table 1 shows the five top-ranked pathways having minimum averaged test error rate in each dataset. The full list of pathway is shown in Supplementary Table 2. Pathways with minimum averaged test error rate of 0.05 and 0.1 were selected to be significant in HCV-positive HCC dataset and HBV-positive HCC dataset, respectively. PAM and GSA algorithms displayed a more similar result in the HBV-positive HCC dataset than in the HCV-positive HCC dataset. For example, in the HBV-positive HCC dataset, many pathways with a false discovery rate (FDR) of 0 by GSA were also selected as significant pathways with the minimum averaged test error rate by PAM. In the HCV-positive HCC dataset, pathways with low averaged error rate in PAM displayed high FDR in GSA. Actually, a large number of pathways selected by GSA displayed poor averaged test error rates, suggesting that PAM is the more suitable method to identify significant pathways. To investigate the distribution of pathways among datasets, we selected the 100 top-ranked pathways having low averaged test error rates by PAM or low FDR by GSA from each dataset. Figs. 3A and B show the Venn diagram for the distribution of these pathways among three datasets. For the PAM algorithm, 11 pathways were commonly selected in all datasets (Fig. 3A), while 10 pathways were common using GSA (Fig. 3B). When we focused on only the HCV-positive dataset and HBV-positive dataset, 23 pathways were common in both types of HCC datasets by the PAM and GSA algorithms. In the HBV- or HCV-positive HCC dataset, which contained HCV-positive samples and HBV-positive samples, more pathways were in common with the HBV-positive HCC dataset (52 pathways by PAM) than the HCV-positive HCC dataset (23 pathways by PAM). This pattern of pathway distribution was also confirmed by GSA. The biological functions of commonly or differently distributed pathways among datasets are shown in Fig. 3C. Intriguingly, cancer-related signaling pathways and lipid metabolism pathways were predominantly enriched in the HCV-positive HCC dataset, whereas diverse pathways including immune-related pathways, cell cycle pathways and RNA metabolism pathways were enriched in the HBV-positive HCC dataset. Some cancer-related signaling pathways and membrane-related maintenance functions were commonly significant in both datasets.

2.3. Genes implicated in pathways

Although we identified commonly selected pathways in the HCV-positive HCC dataset and HBV-positive HCC dataset, this did not indicate the presence of common genes implicated in those pathways.

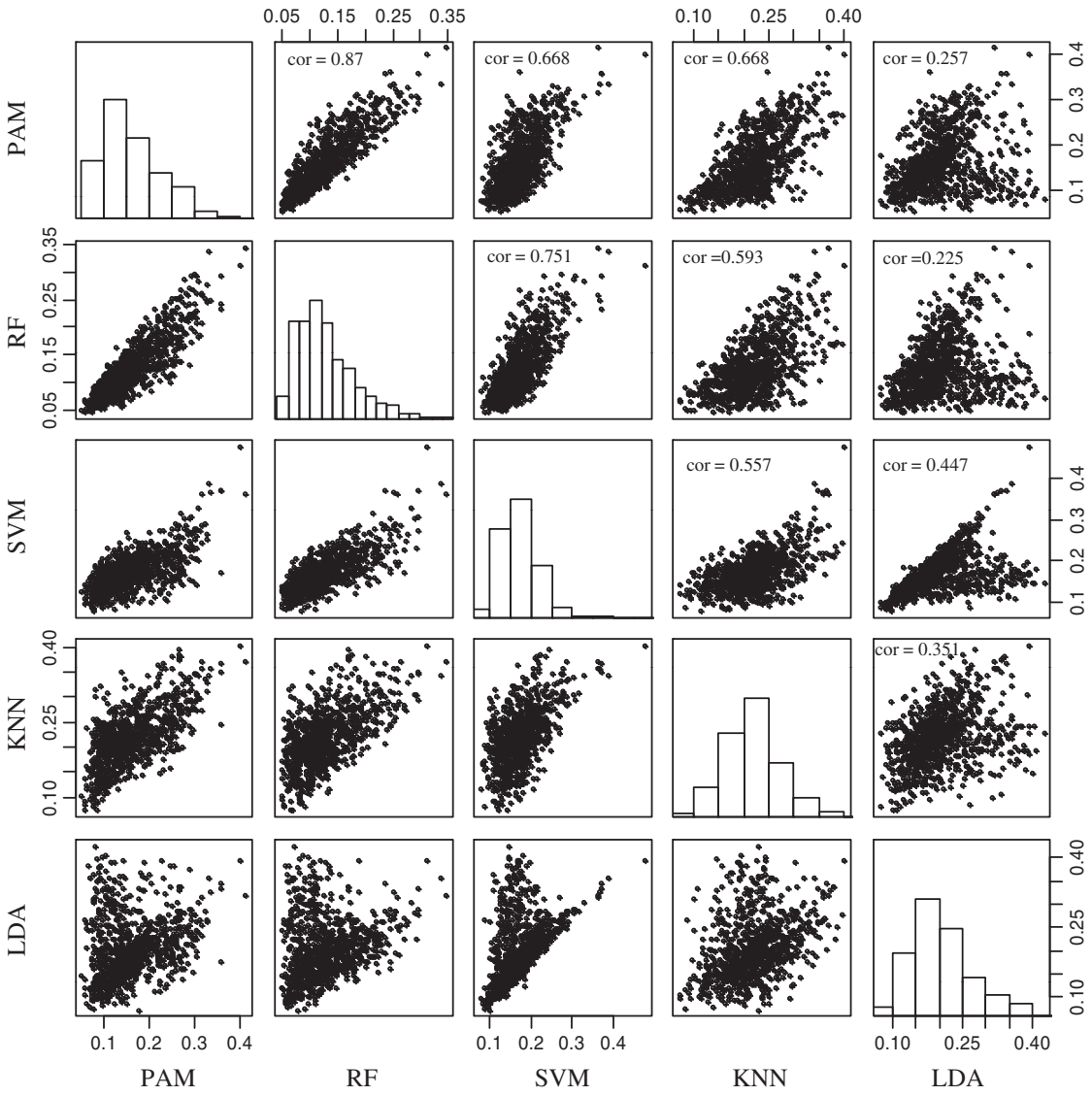

Fig. 2. Scatter plot matrix with histogram of averaged test error on the diagonal and pairwise comparisons of the averaged test error among the five different prediction methods (PAM, RF, SVM, KNN and LDA) on the off-diagonal for the HCV-positive HCC dataset with correlation coefficient.

For example, in mitogen-activated protein kinase (MAPK) signaling pathway, axon guidance pathway and transforming growth factor-beta (TGF-β) signaling pathway, which were all enriched in both the HCV-positive HCC dataset and HBV-positive HCC dataset, the expression pattern of genes was different depending on the dataset (Supplementary Fig. 3). Therefore, the activities of these pathways might be different in each dataset.

2.4. Pathway clusters

We classified pathways into subgroups based on the functional categories in Fig. 3C. Next, pathway clusters were measured based on similarity of class discrimination ability of pathways. Although the overall cluster structure was different between the HCV-positive HCC dataset and HBV-positive HCC dataset, cancer-related signaling pathways and metabolism-related pathways were clustered into separate subgroups as Class 1 and Class 2, respectively, in both datasets ($p<0.01$) (Fig. 4). This result suggests the presence of network of pathways commonly acting on the development of both types of HCC.

3. Discussion

HBV and HCV are completely different viruses. HBV contains a double-stranded DNA genome that integrates into the host genome, while HCV is a RNA virus that replicates in the cytoplasm of the cell [2]. Despite their different life cycles and genomes, they share common characteristics in chronic liver diseases such as hepatitis, which can progress to cirrhosis and HCC. Although this process is not distinguishable by histological examination or clinical manifestations, molecular investigations have identified the differentially expressed genes between HBV- and HCV-positive HCC [5–8]. This result strongly suggests that diagnostic and therapeutic targets for HCC should be considered differently between HBV- and HCV-positive HCC.

In the present study, we measured a pathway's ability to discriminate tumor tissues from adjacent non-tumor liver tissues from the HCV-positive HCC dataset or HBV-positive HCC dataset. When this approach was compared with conventional gene-based approach, it was evident that pathway-based method was less affected by experimental variations from multiple microarray platforms. For example, 23% of pathways were commonly selected between HCV-positive HCC dataset and HBV-positive HCC dataset in our approach, whereas

Table 1
Significant pathways associated with each HCC dataset.

Pathway		PAM				GSA*		
		Error rate**	P**	CV Error***	P***	Score	P	FDR
HCV-positive	Sphingolipid metabolism	0.050	0.054	0.057	0.006	0.253	0.15	0.642
	PDGF pathway	0.057	0.163	0.057	0.028	−0.064	0.33	0.768
	Downstream signal transduction	0.059	0.199	0.046	0.032	0.323	0.08	0.555
	Small cell lung cancer	0.060	0.271	0.057	0.057	0.278	0.05	0.555
	Hedgehog signaling pathway	0.061	0.213	0.057	0.038	−0.367	0.04	0.523
HBV-positive	TGF beta signaling pathway	0.102	0.010	0.094	0.002	−0.667	0.00	0.000
	Mitotic M-M/G1 phases	0.106	0.159	0.099	0.107	0.398	0.00	0.000
	Antigen processing and presentation	0.116	0.020	0.109	0.019	0.804	0.02	0.331
	mRNA splicing	0.119	0.173	0.109	0.125	0.414	0.00	0.000
	Apoptosis	0.121	0.169	0.115	0.133	0.510	0.00	0.000
HBV- or HCV-positive	Metabolism of proteins	0.052	0.139	0.049	0.077	0.472	0.00	0.000
	Influenza life cycle	0.055	0.100	0.043	0.038	0.507	0.00	0.000
	Lysosome	0.056	0.175	0.049	0.123	0.333	0.04	0.276
	Platelet activation	0.058	0.260	0.056	0.266	−0.090	0.17	0.765
	Translation	0.058	0.040	0.056	0.014	0.473	0.01	0.112

* For GSA, score represents gene set score measured by maxmean statistic. P-value and FDR were obtained using restandardization method implemented in GSA R package.
** Error rate was computed by averaging over-all the prediction error rates. P-value was measured by random permutation.
*** CV Error represents 10-fold cross-validated error rate. P-value was measured by random permutation.

only 4.6% of genes were in common in gene-based approach (Fig. 3A and Supplementary Fig. 4). The clinicopathological characteristics also support the similar pathophysiology of three HCC datasets; HCV-positive dataset, HBV-positive dataset and HBV- or HCV-positive dataset (Supplementary Table 1). Variables related with liver pathology; fibrosis stage distribution, total bilirubin, alanine aminotransferase and alpha-fetoprotein level were similar among datasets. Tumor characteristics such as tumor number, size and stage also did not significantly differ among datasets, although the HBV-positive HCC dataset was composed of more poorly differentiated tumor samples (25.8%) compared with other datasets (~18% in HCV-positive HCC dataset and HBV- or HCV-positive HCC dataset, $p = 0.0497$). This clinicopathological information demonstrated that physiological differences among three datasets were mainly derived from viral infection status.

Intriguingly, signaling pathways including diverse cancer pathways were mainly enriched in HCV-positive HCC. In contrast, immune-related pathways, cell cycle pathways and RNA metabolism pathways were mainly enriched in HBV-positive HCC, suggesting the presence of different molecular mechanism in hepatocarcinogenesis

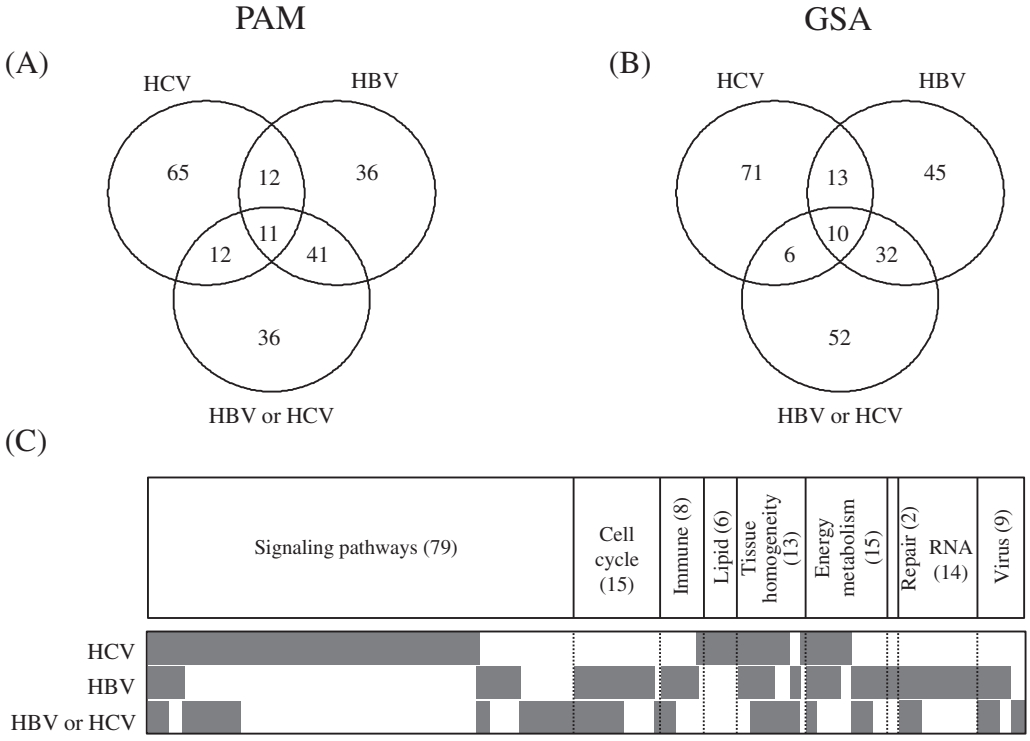

Fig. 3. Comparison of significant pathways among the HCV-positive HCC dataset, HBV-positive HCC dataset and HBV- or HCV-positive HCC dataset. (A) Top ranked 100 pathways with low averaged test error for PAM and (B) top ranked 100 pathways with low FDR for GSA were selected from each dataset. Overlapping of the pathways among datasets was measured in a Venn diagram. (C) Functional distribution of significant pathways (gray color) was measured among datasets. Functional category was divided as follows: cancer/signaling-related class, cell cycle class, immune class, lipid metabolism class, tissue homogeneity class, energy metabolism class, nucleotide repair class, RNA metabolism class and virus-related class. The number in parentheses represents the number of pathways included in each category.

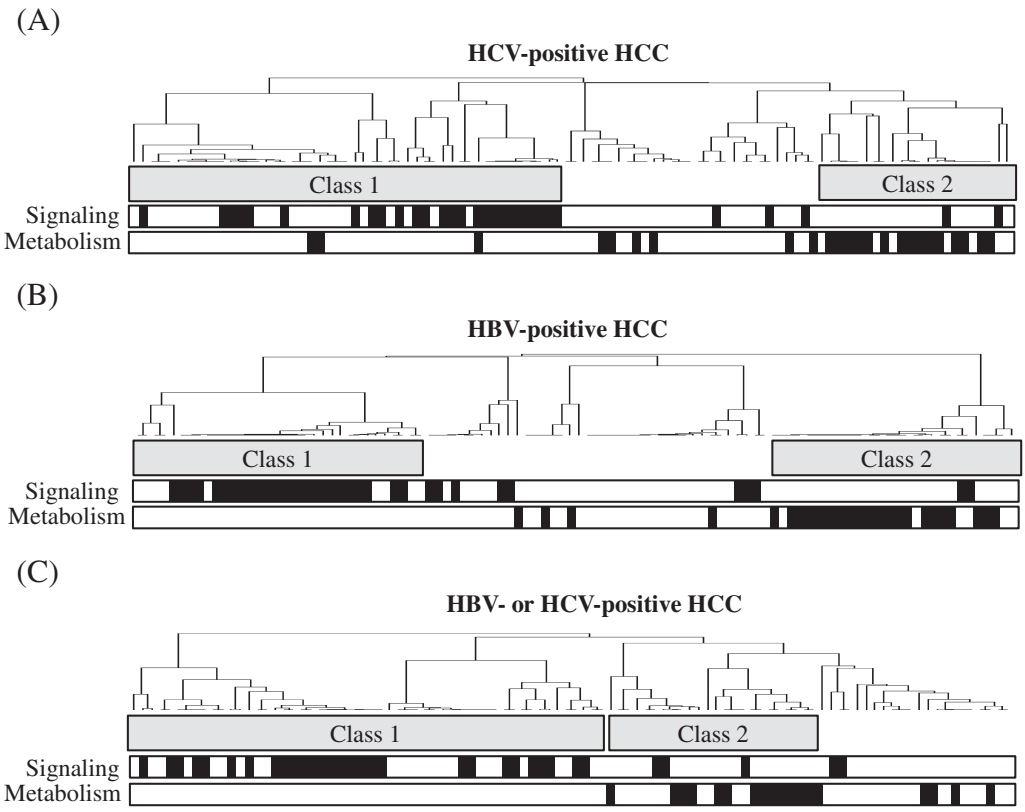

Fig. 4. Tree structure of pathway cluster. Top-ranked 100 pathways were classified into subclasses on similarity of classification error matrix using consensus clustering method for the (A) HCV-positive HCC dataset, (B) HBV-positive HCC dataset and (C) HBV- or HCV-positive HCC dataset. Classes 1 and 2 represent clusters enriched with signaling pathways and metabolism-related pathways, respectively, as colored in black bar.

depending on the status of viral infection. Involvement of signaling pathways such as diverse cancer-related, apoptosis, Wnt and janus kinase/signal transducer and activator of transcription (JAK/STAT) pathways indicates de-regulation of signal pathways in HCV-positive HCC (Table 1 and Supplementary Table 2). It has been reported that HCV infection de-regulates many signal pathways and causes tumor development, although which viral protein plays a key role has not been fully elucidated [22,23]. In addition to cancer- and signal pathways, we identified other functional changes in lipid metabolism in HCV-positive HCC. The glycero-phopholipid metabolism pathway, inositol-phosphate metabolism pathway and sphingolipid metabolism pathway were enriched in HCV-positive HCC. These findings are consistent with the recent report that HCV infection induces abnormality of lipid metabolism and contributes to hepatic steatosis and the development of cancer [24,25]. In addition, the presently-identified autophagy pathway was reported to be critical in suppressing innate antiviral immunity in HCV infection [26].

While signaling pathways were mainly enriched in HCV-positive HCC, relatively diverse biological functions were implicated in HBV-positive HCC suggesting that hepatocarcinogenesis from chronic HBV infection causes more various changes in cellular function than in HCV infection. Although partly involved also in HCV-positive HCC, diverse functional pathways including immune pathways, cell cycle pathways and RNA metabolism pathway were selected as significant pathways in HBV-positive HCC. Previous studies have indicated that dynamic interactions among HBV, hepatocytes and the host immune system may determine viral persistence and disease progression [5,6,27]. Recently, genetic variations at the locus involved in immune response were also reported to be risk factors for HCC [28]. Moreover, involvement of S-phase kinase-associated protein 2 (SKP2)-mediated p21 degradation pathway in HBV-positive HCC, a cell cycle pathway, is consistent with the recent report that mutation in HBV core promoter increases the risk for HCC development by up-regulating SKP2 and then down-regulating p21 via ubiquitin-mediated proteasomal degradation [29]. In addition, enrichment of a group of RNA processing pathways in HBV-positive HCC may also be critical in HBV-positive HCC, as evidenced by the finding that aberrant splicing of mRNA is associated with HCC development and progression [30,31].

These differences between HBV- and HCV-positive HCCs indicate the different molecular mechanism of hepatocarcinogenesis caused by two types of viruses. Although cirrhosis induced by HBV or HCV is a common major risk factor for HCC development, it has been demonstrated that several viral factors including the HBx, pre-S2/S and spliced protein in HBV, and Core, E2 and NS5A in HCV have oncogenic properties acting on different targets in the host [32]. Furthermore, HBV integrated into host genome leads to global changes in genomic function and chromosomal instability.

Previous reports were mainly focused on the identification of difference between HBV- and HCV-positive HCC. However, the present study identified that many cancer-related signaling pathways are commonly significant in both types of HCC, which implies the presence of a common hepatocarcinogenesis process. For example, the TGF-β pathway, MAPK pathway and p53 pathway were included in this category, all of which were already reported to be involved in HCC development and progression [33–35]. Another notable common feature in both HBV-positive HCC and HCV-positive HCC was membrane-related maintenance function including actin cytoskeleton, focal adhesion and axon guidance pathways (Supplementary Table 2). Because maintenance of tissue homeostasis is important in the control of cell growth and differentiation, de-regulated tissue maintenance is critically implicated in tumor progression and metastasis in HCC [36,37]. For example, tight junctions play a key role in HCV entry into host cells [38].

In addition to signaling pathways, energy metabolism pathways including glycolysis have emerged as a potent driving force of liver tumorigenesis [39,40]. We also identified that general metabolic pathways such as glycolysis, oxidative phosphorylation, amino-acid metabolism and nucleotide metabolism were also highly significant in both types of HCC. Considering that Wnt signaling induces a shift in the glucose metabolism from oxidative phosphorylation to glycolysis in the liver [41], the signaling pathways regulating general metabolism may be a key target to control the HCC development.

Although we applied pathway-based class discrimination method to overcome the limitation associated with single gene-based approach, problems involving small size of samples and different types of microarray platforms can still influence on process of the pathway identification. Therefore, it would be important to increase sample numbers with diverse types of microarrays to demonstrate the efficacy of our pathway-based approach in extracting significant biological information.

4. Conclusions

In conclusion, we identified diverse pathways implicated in HCC development according to the status of viral infection. Our findings clearly demonstrate the differences and similarities in biological functions between HBV- and HCV-positive HCC, and the possible presence of a global network of pathways in the development of both types of HCC.

5. Materials and methods

5.1. HCC dataset

Three different microarray datasets of HCC were used in our study. The first set (HCV-positive HCC dataset) was composed of 87 only HCV-positive specimen containing 43 HCCs and 44 non-tumor liver tissues [42] in which, total RNA from frozen samples, or human reference RNA was labeled with fluorescent dyes (Cy5 and Cy3, respectively), and hybridized on arrays (Agilent Technologies). Raw microarray data was archived in Gene Expression Omnibus (GSE17856). For normalization, the log2 values of probe intensity ratio (Cy5/Cy3) were smoothed by LOWESS method [43]. Multiple probes per single gene were averaged and 19,371 genes were finally included for study.

As the second microarray dataset, we used only HBV-positive samples (HBV-positive HCC dataset, GSE14811), previously reported by us [10]. This dataset was composed of 96 HCCs and 96 pair-matched non-tumor liver tissues. Total RNA from each frozen sample and placental reference RNA were labeled with fluorescent dyes (Cy5 and Cy3, respectively) and hybridized with approximately 14,000 cDNAs printed onto glass microscope slides. The log2 ratios of probe intensity (Cy5/Cy3) were normalized using LOWESS method. Especially, space- and intensity-dependent LOWESS method was applied to eliminate intensity bias associated with manufacturing process of spot-type slide microarray in this dataset [44]. After averaging ratios of multiple probes per single gene, 6122 genes were included in the present study.

Finally, the third dataset (HBV- or HCV-positive HCC dataset, GSE10143) included 162 samples (80 HCCs and 82 non-tumor liver tissues) composed of heterogeneous viral types of specimen from Hosida et al. [45]. The samples had been kept in formalin-fixed and paraffin-embedded blocks. Total RNA extracted from tissues was converted into cDNA and then was employed to the cDNA-mediated annealing, selection, extension, and ligation (DASL) assay (Illumina). The amplified products were hybridized to a bead microarray. The one color signal intensities of quality controlled 6100 genes were normalized using quantile method to make the distribution of probe intensities of each array the same [43].

5.2. Classification algorithm

Fig. 1 shows a schematic diagram of overall procedure applied to identify significant pathways. Initially, each microarray dataset was split into two groups randomly as the training and test set composed of 60% and 40% of the samples, respectively. With pre-defined genes belonging to each pathway, the prediction model was built by implementing five different algorithms in the training set to discriminate tumor samples from adjacent non-tumor samples; nearest shrunken centroid (or Prediction Analysis of Microarrays; PAM) [21], random forest (RF) [19], support vector machine (SVM) [46], linear discriminant analysis (LDA), and k-nearest neighbor classifier (KNN). Finally, the prediction error rate was evaluated on independent test data. This procedure was repeated 1000 times to extract as much information as possible from all samples. Then, the mean test error rate was computed by averaging over-all the prediction error rates for a pathway. Finally, to estimate the statistical significance of an averaged test error rate, the permutation-based approach was used. The gene labels (gene symbol) were randomly permuted 1000 times. For each permutation, the random test error rate was measured as the same procedure used in the original dataset. By comparing the original averaged test error rate with permutated random error rates, the statistical significance for a pathway was estimated. We applied a gene-based permutation rather than a class (phenotype)-based permutation, reflecting the fact that the expression of many genes are already changed between HCC and adjacent non-tumor liver tissues [9,10].

We also measured the pathway prediction efficacy using a cross-validation (CV) method. In that analysis, the prediction model was built and evaluated on 10-fold CV for PAM. The statistical significance of cross-validated error rate was also measured using gene-based random permutation method ($n = 1000$). All procedures were performed using R (v2.12.0; the R source code for our program is available upon request). On the other hand, *t*-test was used for selection of genes discriminating between HCC and adjacent non-tumor liver tissue.

5.3. Gene set analysis (GSA)

GSA is an improved version of Gene Set Enrichment Analysis (GSEA) [11,18]. GSA measures the gene-set score for each gene set and searches for significantly correlated gene sets with the phenotypic class. For comparison with our method, we performed GSA algorithm using GSA R package (v1.03).

5.4. Pathway information

The pathway database was obtained from Molecular Signatures Database (MSigDB) [11], from which manually curated pathway information of Kyoto Encyclopedia of Genes and Genomes (KEGG, 186 pathways) [47], BioCarta (217 pathways, http://www.biocarta.com/genes/allpathways.asp) and Reactome database (430 pathways) [48] were initially included in this study. From a total of 833 pathways, only pathways containing expression values of at least five genes were included in this study. Therefore, 818 pathways (185 from KEGG, 214 from BioCarta and 419 from Reactome) for HCV-positive HCC dataset, 680 pathways (175 from KEGG, 159 from BioCarta and 346 from Reactome) for HBV-positive HCC dataset and 799 pathways (183 from KEGG, 216 from BioCarta and 400 from Reactome) for HBV- or HCV-positive HCC dataset were used.

5.5. Pathway cluster

The similarity of pathways was measured by the consensus clustering method [49], which is a tool for unsupervised class discovery involving subsampling. The matrix of classification error rate measured on each sample and each pathway using PAM algorithms was used for clustering (ConsensusClusterPlus R package v1.0.1). We included 100 top-ranked pathways from KEGG having low averaged test error. Cluster count (k) of 5 was applied after graphical determination of cumulative distribution function. Statistical significance for functional clustering of pathways was measured by Chi-square test performed in R.

Acknowledgments

This study was supported by a grant (kiom-2010-2) from the Inter-Institutional Collaboration Research Program under the Korea Research Council of Fundamental Science & Technology (KRCF) and by the National Research Foundation of Korea grant (NRF, No. 20110027738).

Appendix A. Supplementary data

Supplementary data to this article can be found online at doi:10.1016/j.ygeno.2012.04.004.

References

1 D.M. Parkin, F. Bray, J. Ferlay, P. Pisani, Global cancer statistics, 2002, CA Cancer J. Clin. 55 (2005) 74–108.
2 R. Colombari, A.P. Dhillon, E. Piazzola, A.A. Tomezzoli, G.P. Angelini, F. Capra, A. Tomba, P.J. Scheuer, Chronic hepatitis in multiple virus infection: histopathological evaluation, Histopathology 22 (1993) 319–325.
3 M. Honda, S. Kaneko, H. Kawai, Y. Shirota, K. Kobayashi, Differential gene expression between chronic hepatitis B and C hepatic lesion, Gastroenterology 120 (2000) 955–966.
4 M. Honda, T. Yamashita, T. Ueda, H. Takatori, R. Nishino, S. Kaneko, Different signaling pathways in the livers of patients with chronic hepatitis B or chronic hepatitis C, Hepatology 44 (2006) 1122–1138.
5 N. Iizuka, M. Oka, H. Yamada-Okabe, N. Mori, T. Tamesa, T. Okada, N. Takemoto, A. Tangoku, K. Hamada, H. Nakayama, T. Miyamoto, S. Uchimura, Y. Hamamoto, Comparison of gene expression profiles between hepatitis B virus- and hepatitis C virus-infected hepatocellular carcinoma by oligonucleotide microarray data on the basis of a supervised learning method, Cancer Res. 62 (2002) 3939–3944.
6 N. Iizuka, M. Oka, H. Yamada-Okabe, N. Mori, T. Tamesa, T. Okada, N. Takemoto, K. Hashimoto, A. Tangoku, K. Hamada, H. Nakayama, T. Miyamoto, S. Uchimura, Y. Hamamoto, Differential gene expression in distinct virologic types of hepatocellular carcinoma: association with liver cirrhosis, Oncogene 22 (2003) 3007–3014.
7 C.F. Lee, Z.Q. Ling, T. Zhao, K.R. Lee, Distinct expression patterns in hepatitis B virus- and hepatitis C virus-infected hepatocellular carcinoma, World J. Gastroenterol. 14 (2008) 6072–6077.
8 S.Y. Yoon, J.M. Kim, J.H. Oh, Y.J. Jeon, D.S. Lee, J.H. Kim, J.Y. Choi, B.M. Ahn, S. Kim, H.S. Yoo, Y.S. Kim, N.S. Kim, Gene expression profiling of human HBV- and/or HCV-associated hepatocellular carcinoma cells using expressed sequence tags, Int. J. Oncol. 29 (2006) 315–327.
9 B.Y. Kim, J.G. Lee, S. Park, J.Y. Ahn, Y.J. Ju, J.H. Chung, C.J. Han, S.H. Jeong, Y.I. Yeom, S. Kim, Y.S. Lee, C.M. Kim, E.M. Eom, D.H. Lee, K.Y. Choi, M.H. Cho, K.S. Suh, D.W. Choi, K.H. Lee, Feature genes of hepatitis B virus-positive hepatocellular carcinoma, established by its molecular discrimination approach using prediction analysis of microarray, Biochim. Biophys. Acta 1739 (2004) 50–61.
10 S.H. Chang, K.S. Suh, N.J. Yi, K.H. Lee, B.Y. Kim, J.J. Jang, Predicting the prognosis of hepatocellular carcinoma using gene expression, J. Surg. Res. 171 (2011) 524–531.
11 A. Subramanian, P. Tamayo, V.K. Mootha, S. Mukherjee, B.L. Ebert, M.A. Gillette, A. Paulovich, S.L. Pomeroy, T.R. Golub, E.S. Lander, J.P. Mesirov, Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles, Proc. Natl. Acad. Sci. U. S. A. 102 (2005) 15545–15550.
12 L. Abatangelo, R. Maglietta, A. Distaso, A. D'Addabbo, T.M. Creanza, S. Mukherjee, N. Ancona, Comparative study of gene set enrichment methods, BMC Bioinformatics 10 (2009) 275.
13 M. Murohashi, K. Hinohara, M. Kuroda, T. Isagawa, S. Tsuji, S. Kobayashi, K. Umezawa, A. Tojo, H. Aburatani, N. Gotoh, Gene set enrichment analysis provides insight into novel signalling pathways in breast cancer stem cells, Br. J. Cancer 102 (2010) 206–212.
14 N.G. Copeland, N.A. Jenkins, Deciphering the genetic landscape of cancer—from genes to pathways, Trends Genet. 25 (2009) 455–462.
15 M.A. Ali, T. Sjoblom, Molecular pathways in tumor progression: from discovery to functional understanding, Mol. Biosyst. 5 (2009) 902–908.
16 F. Tai, W. Pan, Incorporating prior knowledge of predictors into penalized classifiers with multiple penalty terms, Bioinformatics 23 (2007) 1775–1782.
17 M.C. Wu, L. Zhang, Z. Wang, D.C. Christiani, X. Lin, Sparse linear discriminant analysis for simultaneous testing for the significance of a gene set/pathway and gene selection, Bioinformatics 25 (2009) 1145–1151.
18 B. Efron, R. Tibshirani, On testing the significance of sets of genes, Ann. Appl. Stat. 1 (2007) 107–129.
19 H. Pang, A. Lin, M. Holford, B.E. Enerson, B. Lu, M.P. Lawton, E. Floyd, H. Zhao, Pathway analysis using random forests classification and regression, Bioinformatics 22 (2006) 2028–2036.
20 H. Pang, H. Zhao, Building pathway clusters from Random Forests classification using class votes, BMC Bioinformatics 9 (2008) 87.
21 R. Tibshirani, T. Hastie, B. Narasimhan, G. Chu, Diagnosis of multiple cancer types by shrunken centroids of gene expression, Proc. Natl. Acad. Sci. U. S. A. 99 (2002) 6567–6572.
22 M. Levrero, Viral hepatitis and liver cancer: the case of hepatitis C, Oncogene 25 (2006) 3834–3847.
23 K. Koike, Hepatitis C virus contributes to hepatocarcinogenesis by modulating metabolic and intracellular signaling pathways, J. Gastroenterol. Hepatol. 22 (Suppl. 1) (2007) S108–S111.
24 J.M. Wu, N.J. Skill, M.A. Maluccio, Evidence of aberrant lipid metabolism in hepatitis C and hepatocellular carcinoma, HPB (Oxford) 12 (2010) 625–636.
25 K. Koike, Steatosis, liver injury, and hepatocarcinogenesis in hepatitis C viral infection, J. Gastroenterol. 44 (Suppl. 19) (2009) 82–88.
26 P.Y. Ke, S.S. Chen, Activation of the unfolded protein response and autophagy after hepatitis C virus infection suppresses innate antiviral immunity in vitro, J. Clin. Invest. 121 (2011) 37–56.
27 T.F. Baumert, R. Thimme, F. von Weizsacker, Pathogenesis of hepatitis B virus infection, World J. Gastroenterol. 13 (2007) 82–90.
28 R.J. Clifford, J. Zhang, D.M. Meerzaman, M.S. Lyu, Y. Hu, C.M. Cultraro, R.P. Finney, J.M. Kelley, S. Efroni, S.I. Greenblum, C.V. Nguyen, W.L. Rowe, S. Sharma, G. Wu, C. Yan, H. Zhang, Y.H. Chung, J.A. Kim, N.H. Park, I.H. Song, K.H. Buetow, Genetic variations at loci involved in the immune response are risk factors for hepatocellular carcinoma, Hepatology 52 (2010) 2034–2043.
29 Y. Huang, S. Tong, A.W. Tai, M. Hussain, A.S. Lok, Hepatitis B virus core promoter mutations contribute to hepatocarcinogenesis by deregulating SKP2 and its target, p21, Gastroenterology 141 (2011) 1412–1421.
30 X. Lu, X. Feng, X. Man, G. Yang, L. Tang, D. Du, F. Zhang, H. Yuan, Q. Huang, Z. Zhang, Y. Liu, D. Strand, Z. Chen, Aberrant splicing of Hugl-1 is associated with hepatocellular carcinoma progression, Clin. Cancer Res. 15 (2009) 3287–3796.
31 X.Q. Wang, J.M. Luk, P.P. Leung, B.W. Wong, E.J. Stanbridge, S.T. Fan, Alternative mRNA splicing of liver intestine-cadherin in hepatocellular carcinoma, Clin. Cancer Res. 11 (2005) 483–489.
32 J. Fung, C.L. Lai, M.F. Yuen, Hepatitis B and C virus-related carcinogenesis, Clin. Microbiol. Infect. 15 (2009) 964–970.
33 G. Giannelli, A. Mazzocca, E. Fransvea, M. Lahn, S. Antonaci, Inhibiting TGF-β signaling in hepatocellular carcinoma, Biochim. Biophys. Acta 1815 (2011) 214–223.
34 H.J. Baek, M.J. Pishvaian, Y. Tang, T.H. Kim, S. Yang, M.E. Zouhairi, J. Mendelson, K. Shetty, B. Kallakury, D.L. Berry, K.H. Shin, B. Mishra, E.P. Reddy, S.S. Kim, L. Mishra, Transforming growth factor-beta adaptor, beta2-spectrin, modulates cyclin dependent kinase 4 to reduce development of hepatocellular cancer, Hepatology 53 (2011) 1676–1684.
35 L. Min, B. He, L. Hui, Mitogen-activated protein kinases in hepatocellular carcinoma development, Semin. Cancer Biol. 21 (2011) 10–20.
36 Y. Nakashima, T. Ono, A. Yamanoi, O.N. El-Assal, H. Kohno, N. Nagasue, Expression of gap junction protein connexin32 in chronic hepatitis, liver cirrhosis, and hepatocellular carcinoma, J. Gastroenterol. 39 (2004) 763–768.
37 T. Sakaguchi, S. Suzuki, H. Higashi, K. Inaba, S. Nakamura, S. Baba, T. Kato, H. Konno, Expression of tight junction protein claudin-5 in tumor vessels and sinusoidal endothelium in patients with hepatocellular carcinoma, J. Surg. Res. 147 (2008) 123–131.
38 M.B. Zeisel, I. Fofana, S. Fafi-Kremer, T.F. Baumert, Hepatitis C virus entry into hepatocytes: molecular mechanisms and targets for antiviral therapies, J. Hepatol. 54 (2011) 566–576.
39 T. Amann, U. Maegdefrau, A. Hartmann, A. Agaimy, J. Marienhagen, T.S. Weiss, O. Stoeltzing, C. Warnecke, J. Scholmerich, P.J. Oefner, M. Kreutz, A.K. Bosserhoff, C. Hellerbrand, GLUT1 expression is increased in hepatocellular carcinoma and promotes tumorigenesis, Am. J. Pathol. 174 (2009) 1544–1552.
40 K. Daskalow, D. Pfander, W. Weichert, N. Rohwer, A. Thelen, P. Neuhaus, S. Jonas, B. Wiedenmann, C. Benckert, T. Cramer, Distinct temporospatial expression patterns of glycolysis-related proteins in human hepatocellular carcinoma, Histochem. Cell Biol. 132 (2009) 21–31.
41 P. Chafey, L. Finzi, R. Boisgard, M. Cauzac, G. Clary, C. Broussard, J.P. Pegorier, F. Guillonneau, P. Mayeux, L. Camoin, B. Tavitian, S. Colnot, C. Perret, Proteomic analysis of beta-catenin activation in mouse liver by DIGE analysis identifies glucose metabolism as a new target of the Wnt pathway, Proteomics 9 (2009) 3889–3900.
42 M. Tsuchiya, J.S. Parker, H. Kono, M. Matsuda, H. Fujii, I. Rusyn, Gene expression in nontumoral liver tissue and recurrence-free survival in hepatitis C virus-positive hepatocellular carcinoma, Mol. Cancer 9 (2010) 74.
43 B.M. Bolstad, R.A. Irizarry, M. Astrand, T.P. Speed, A comparison of normalization methods for high density oligonucleotide array data based on variance and bias, Bioinformatics 19 (2003) 185–193.
44 Y.H. Yang, S. Dudoit, P. Luu, D.M. Lin, V. Peng, J. Ngai, T.P. Speed, Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation, Nucleic Acids Res. 30 (2002) e15.
45 Y. Hoshida, A. Villanueva, M. Kobayashi, J. Peix, D.Y. Chiang, A. Camargo, S. Gupta, J. Moore, M.J. Wrobel, J. Lerner, M. Reich, J.A. Chan, J.N. Glickman, K. Ikeda, M. Hashimoto, G. Watanabe, M.G. Daidone, S. Roayaie, M. Schwartz, S. Thung, H.B.

Salvesen, S. Gabriel, V. Mazzaferro, J. Bruix, S.L. Friedman, H. Kumada, J.M. Llovet, T.R. Golub, Gene expression in fixed tissues and outcome in hepatocellular carcinoma, N. Engl. J. Med. 359 (2008) 1995–2004.
46 C.C. Chang, C.J. Lin, LIBSVM: a library for support vector machines, ACM TIST 2 (2011) 27.
47 M. Kanehisa, M. Araki, S. Goto, M. Hattori, M. Hirakawa, M. Itoh, T. Katayama, S. Kawashima, S. Okuda, T. Tokimatsu, Y. Yamanishi, KEGG for linking genomes to life and the environment, Nucleic Acids Res. 36 (2008) D480–D484.
48 L. Matthews, G. Gopinath, M. Gillespie, M. Caudy, D. Croft, B. de Bono, P. Garapati, J. Hemish, H. Hermjakob, B. Jassal, A. Kanapin, S. Lewis, S. Mahajan, B. May, E. Schmidt, I. Vastrik, G. Wu, E. Birney, L. Stein, P. D'Eustachio, Reactome knowledgebase of human biological pathways and processes, Nucleic Acids Res. 37 (2009) D619–D622.
49 S. Monti, P. Tamayo, J. Mesirov, T. Golub, Consensus clustering: a resampling-based method for class discovery and visualization of gene expression microarray data, Mach. Learn. 52 (2003) 91–118.